

Statistical validation of critical aspects of the Net Promoter Score

Validation of
NPS's critical
aspects

Manuela Cazzaro and Paola Maddalena Chiodini
University of Milan-Bicocca, Milan, Italy

191

Abstract

Purpose – Although the Net Promoter Score (NPS) index is simple, NPS has weaknesses that make NPS's interpretation misleading. The main criticism is that identical index values can correspond to different levels of customer loyalty. This makes difficult to determine whether the company is improving/deteriorating in two different years. The authors describe the application of statistical tools to establish whether identical values may/may not be considered similar under statistical hypotheses.

Design/methodology/approach – Equal NPSs with a “similar” component composition should have a two-way table satisfying marginal homogeneity hypothesis. The authors compare the marginals using a cumulative marginal logit model that assumes a proportional odds structure: the model has the same effect for each logit. Marginal homogeneity corresponds to null effect. If the marginal homogeneity hypothesis is rejected, the cumulative odds ratio becomes a tool for measuring the proportionality between the odds.

Findings – The authors propose an algorithm that helps managers in their decision-making process. The authors' methodology provides a statistical tool to recognize customer base compositions. The authors suggest a statistical test of the marginal distribution homogeneity of the table representing the index compositions at two times. Through the calculation of cumulative odds ratios, the authors discriminate against the hypothesis of equality of the NPS.

Originality/value – The authors' contribution provides a statistical alternative that can be easily implemented by business operators to fill the known shortcomings of the index in the customer satisfaction's context. This paper confirms that although a single number summarizes and communicates a complex situation very quickly, the number is ambiguous and unreliable if not accompanied by other tools.

Keywords Customer satisfaction, NPS, Marginal homogeneity, Cumulative logit model, Cumulative odds ratio

Paper type Case study

Received 9 June 2022
Revised 22 September 2022
13 December 2022
7 March 2023
Accepted 17 March 2023

1. Introduction

Customer satisfaction and retention are very important factors for companies that work in increasingly competitive markets. Following [Arora and Narula \(2018\)](#), “Customer satisfaction is mainly derived from the physiological response with the perceptual difference gap between expectation before consumption and practical experience after consumption of service or products. It implies an accumulated temporary and sensory response.”

The literature is full of proposals for methods to measure customer satisfaction; see, among others, [Ngo \(2015\)](#). Measurement can be approached through the use of various models and methods, of which the best known are Net Promoter Score (NPS), National Customer Satisfaction Index (NCSI), American Customer Satisfaction Index (ACSI), European Performance Satisfaction Index (EPSI), Service Quality (SERVQUAL), probit/logit model, Multicriteria Satisfaction Analysis (MUSA) and statistical regression models based on latent variables. Note that many of these approaches may also involve the use of articulated questionnaires.

© Manuela Cazzaro and Paola Maddalena Chiodini. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

Conflict of interest: The authors declare that they have no conflict of interest.



The TQM Journal
Vol. 35 No. 9, 2023
pp. 191-209
Emerald Publishing Limited
1754-2731
DOI 10.1108/TQM-05-2022-0170

In governance and marketing processes aimed at maximizing a company's success, customer loyalty is of paramount importance, a process that is closely linked to customer satisfaction. In fact, these processes have an impact on satisfaction, and satisfied customers become loyal ones (Arora and Narula, 2018). Measuring the level of satisfaction of a customer with statistical models can be very complex and difficult (Zanella, 1998; De Luca, 2006). The models normally used may not be easy to implement. The variables that govern the mechanisms of customer choice and satisfaction are generally very difficult to measure and model.

Furthermore, the quest to consolidate the company's position in the market and win more market share cannot be separated from the need to understand what the customers want. Their needs change over time, as do their requirements, and this pushes the companies toward a continuous search for improvement as indicated by the philosophy of Total Quality Management (TQM). TQM is a quality-based strategic tool of management and characterizes the basis for successful organization that ensures the success of organizations in the competitive economy. If TQM is effectively evidenced in the quality of the product, customer loyalty is automatically enhanced, Worlu *et al.* (2019). Deming (1986) perceives TQM as a set of management practices that enable companies to increase their productivity and quality by having the ability to create constancy of purpose for improving products and services and stop reliance on inspection to attain quality. The Plan-Do-Check-Act (PDCA), also known as the "Deming wheel", had its origin with Deming's lecture in Japan in 1950 by modifying the Shewhart cycle introduced in 1939. The PDCA cycle (Figure 1) is a widely utilized management methodology in those companies aiming at continuous improvement.

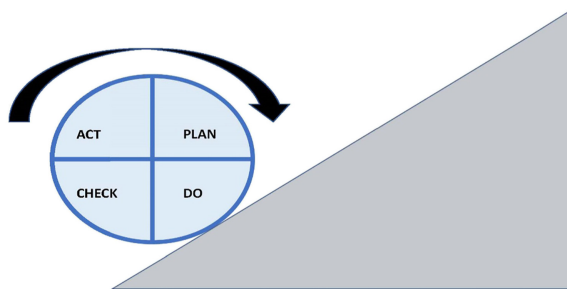


Figure 1.
PDCA cycle

Source(s): Figure by authors

In this context, the customer satisfaction methodologies already indicated also fit in, as the NPS does. The NPS, introduced by Reichheld (2003) and then revised by himself (2011), fits in as a new resource that is agile to use (it is based on a single question) and, above all, that leverages the word-of-mouth (WOM). Loyalty is reflected when customers say positive things about the firm, intend to do business with the company and consider that particular company their first choice. In an increasingly globalized world where e-commerce is expanding rapidly, WOM seems to be a winning aspect for companies that increasingly rely on asking their customers and buyers for ratings to be published online.

The basis of the NPS is the idea that a satisfied customer would be willing to recommend the brand to friends and acquaintances. Reichheld believes that WOM recommendations are a useful, powerful and simple tool for measuring the degree of success of a brand and the degree of its customer loyalty.

The customer is asked a single question: "How likely is it that you would recommend us to a friend or colleague?" The response uses a scale of 11 points, from 0 (indicating "I probably

won't recommend it") up to 10 (indicating "I will most likely recommend it"). The NPS takes into account the responses to this single question. In fact, Reichheld maintains that a higher level of customer satisfaction, and consequently loyalty, will result in a higher score in response to the question.

The scale is divided into three clusters: scores of 9 and 10 indicate clients considered *promoters*, scores of 7 or 8 are considered *neutral* or *passive* clients and scores of 6 or lower are considered *detractors* (see Figure 2).



Source(s): Figure by authors

Figure 2. Client categories used in evaluating NPS

Of the three groups of scores identified, only two are used to calculate the NPS:

$$NPS = \frac{\# Promoters - \# Detractors}{\# Respondents}$$

The NPS measure theoretically ranges from -1 (no promoters and all respondents are detractors) to +1 (all respondents are promoters), although typical values are in the range 0.3-0.4. Obviously, the value can be read as a percentage.

The simple nature of the NPS index has made it very popular and widely used, but it has also generated considerable disagreement. It is clear that this measure has pros and cons.

This paper does not have the ambition to provide a new tool that can replace the well-known NPS, but it instead focuses attention on the indiscriminate use of the score. The aim of the work is to present a statistical methodology already known in the literature (the marginal homogeneity model and the cumulative odds ratio) that, when combined with the NPS index, allows a correct reading of its value. Although the proposed method does not correct the known structural weaknesses of the index, it allows us to begin to answer some of the criticisms raised by allowing an objective reading (see Subsection 1.1).

1.1 Net Promoter Score critical issues

The introduction of the NPS index, in spite of considerable criticism in the scientific community, turns out to be a tool that is easy to implement even by those without specific statistical knowledge. For business operators, the evaluation of the number of potentially satisfied customers (promoters) is easy. Their satisfaction is measured indirectly through the score they give to the possibility of suggesting the brand/product to other possible buyers. This mechanism is believed to trigger a growth/decline process of the company's image on the market with the consequent increase/decline of customers.

Reading this index should help the company to understand not only its position in the market relative to its competitors, but also whether its position has been improving (detractors or passives becoming promoters) or worsening (promoters moving to the position of detractors or passives). Extensive debate has been conducted in the literature regarding the fact that the so-called move from one "state" to the next may not be easy to detect, i.e. the indicator does not provide any insight into the decision-making process or the motivation for the customer to move from one state to the next. Ultimately, there is little doubt that a detractor is unlikely to become a promoter. It is more reasonable to expect that it is the passives (ignored in the calculation of the NPS) who can change the state by altering the state of affairs and the value of the index.

Apple, Amazon, American Express, Avis, HP, Sky and IBM are among the many prominent adopters of NPS. The benchmark is popular for its simplicity, and Reichheld claims it correlates to company growth. Critics contend that this is not the case ([Sharp, 2006](#); [Pingitore et al., 2007](#); [East et al., 2011](#); [Eskildsen and Kristensen, 2011](#); [Kristensen and Eskildsen, 2014](#)). In particular, the 11-point scale is argued to have lower predictive validity than other scales ([Schneider et al., 2008](#)), the segmentation of promoters/passives/detractors is arbitrary and other questions may be better predictors of growth rates as reported by Jeff Sauro [[1](#)] and Richard Evensen [[2](#)] in their blogs:

- (1) The single question is not the most important in terms of customer satisfaction: this means that the NPS is surely less accurate than composite customer satisfaction indices based on, for example, three questions;
- (2) The NPS does not accurately differentiate promoters and detractors: the composition of the three classes proposed by Reichheld is not supported statistically;
- (3) The NPS fails to predict loyalty behavior;
- (4) The NPS performs worse than satisfaction and liking questions;
- (5) The NPS performs worse than other scales;
- (6) The scoring inflates the margin of error: by converting an 11-point scale into a 2-point scale of detractors and promoters, information is lost. Throwing out the “passive” clients means that the organization misses the opportunity to work on those customers that are easiest to move upward to promoters.

Despite enduring managerial popularity, academics remain skeptical of NPS, citing methodological issues and ongoing concerns with NPS measurement. In particular, [Eskildsen and Kristensen \(2011\)](#) and [Kristensen and Eskildsen \(2014\)](#) believe that NPS is not a reliable indicator of effective customer retention. The ability of the NPS to really measure customer satisfaction and, consequently, loyalty to the brand is increasingly being questioned. In fact, there is no evidence linking the growth/decrease of the index to an equivalent growth/decrease in the business volume of the company. The single question used to compute the NPS does not consider the psychological variables that lead to the purchase and repurchase of a specific product/service. Indeed, the consumers who buy durable goods exhibit different behavior from those who buy consumer goods. There is no focus on the customer’s intention to eventually buy the product again in the future, only on his/her propensity to suggest the brand to friends or acquaintances. [Mecredy et al. \(2018\)](#) and [Baehre et al. \(2022\)](#) revisited the use of NPS as a predictor of short-term sales growth through empirical investigations, concluding that the methodological concerns raised by academics are valid. Furthermore, there are considerable differences in the different markets where companies operate. Likewise, the socioeconomic variables used to describe customers are not taken into account. There is also a complete absence of the “do not know” mode in the scale of possible answers, which removes the potential for the respondent to express neutrality.

Companies operating in different markets and having to deal with different dynamics cannot readily compare themselves using the NPS index. Similar values of the NPS index for companies operating in different markets could have completely different meanings in terms of affirmation and the acquisition of market share. However, it is not even clear how one can think of comparing the value of the NPS index between companies operating in the same market. If a company has a higher NPS index value, how should this result be interpreted? And if the value of the index for a given company increases over time, does that mean that the market position is being consolidated and consequently that profits will increase? These multiple aspects are not taken into account in the structure of the NPS. A further criticism of comparing the NPS of similar companies operating in different countries is that some

countries are more accustomed to using the full scale of marks from 0 to 10 following habits formed at school. Nations such as Great Britain or the Scandinavian countries label scores of 9 and 10 as “excellent” due to their cultural heritage; Italian high schools, on the other hand, follow the standard that a mark of 8 out of 10 is considered an “excellent” grade.

[Kristensen and Eskildsen \(2014\)](#) suggested a different distribution of respondents, such that scores from 0 to 4 are attributed to detractors, scores from 5 to 7 are passives and scores from 8 to 10 indicate promoters. This type of clustering will distribute the interviewees in more homogeneous groups. Note that the most well-known and accredited customer satisfaction measurement indicators, such as the EPSI rating or ACSI, use a 10-point score scale, which is considered more efficient.

Given the structure of the NPS index, it is not even clear how the scores should be interpreted. If developed for all possible combinations of percentages in the three clusters, a perfectly symmetrical triangular structure is obtained. This suggests that identical NPS values can be obtained with profoundly different compositions of the percentages involved in the calculation. How should this result be interpreted? Can identical index values indicate the same business performance even if the results derive from different percentage compositions of detractors and promoters?

In their critical review of the NPS, [Fisher and Kordupleski \(2019\)](#) highlight five further problems with as the index:

- (1) The NPS provides no data on how a company can improve;
- (2) The NPS focuses only on keeping customers, not on winning new customers;
- (3) There is no such thing as a “passive” customer;
- (4) The NPS provides no competitive data;
- (5) The NPS is internally focused, not externally focused.

They also provide recommendations on how to avoid these problems.

Despite these criticisms, the NPS remains popular because it is well marketed, easy to understand and its model makes intuitive sense: every organization wants more promoters than detractors.

In this paper, we describe the application of statistical tools to the NPS to establish whether identical values of the NPS index produced by different compositions of customers may or may not be considered similar under statistical hypotheses.

The remainder of this paper is organized as follows. [Section 2](#) recalls some statistical aspects of the NPS already reported in the literature. In particular, we report the proposal presented by [Rocks \(2016\)](#) regarding the confidence interval of the index and the research of [Capecchi and Piccolo \(2017\)](#) on the distribution of NPS. In [Section 3](#), we present a critical comparison of similar NPSs. Specifically, in [Subsection 3.1](#), we analyze equal scores at two different points in time, but referring to indices generated with different compositions. [Section 4](#) presents the methodology we use to introduce our proposal. The *marginal homogeneity test* described in [Subsection 4.1](#) provides a statistical validation of equal NPSs at two different points in time. [Subsection 4.2](#) goes further and suggests that the *cumulative odds ratio* should be adopted to establish the proportionality between the odds of different outcomes. Results are reported in [Section 5](#). Finally, [Section 6](#) presents the discussion, the implications and further research.

2. Statistical aspects of the NPS

The simplicity of the NPS means that it is widely used, despite being heavily criticized. However, only a few recent papers faced inferential procedures with regards to NPS. In particular, [Rocks \(2016\)](#) and [Capecchi and Piccolo \(2017\)](#).

Rocks (2016) describes the properties of the NPS starting from the definition of its distribution law. The goal is to compare different confidence intervals. The main difficulty relates to the definition of the NPS distribution, as many trinomial laws can be suitably adapted. To calculate the variance of the NPS, σ_{NPS} , it seems appropriate to use the difference of two proportions (Gold, 1963; Goodman, 1965), giving the following formula:

$$\sigma_{NPS} = p_{pro} + p_{det} - (p_{pro} - p_{det})^2,$$

where p_{pro} and p_{det} represent the proportions of promoters and detractors, respectively. Different approaches for determining the confidence interval for the NPS were presented by Rocks. Among these, Wald's confidence interval, which is based on Laplace's proposal (de Laplace, 1812), stands out:

$$NPS \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_{NPS}}{n}}$$

where $z_{\alpha/2}$ is the standard normal distribution quantile and n is the sample size. An alternative proposal is the adjusted Wald interval introduced by Agresti and Coull (1998) and subsequently modified by Agresti and Min (2005) for matched pairs in a 2×2 contingency table:

$$\widehat{NPS} \pm z_{\alpha/2} \cdot \sqrt{\frac{\widehat{\sigma}_{NPS}}{\widehat{n}}}$$

where \widehat{NPS} , $\widehat{\sigma}_{NPS}$ and \widehat{n} are the adjusted estimates. Analogous to the Wald method is the Goodman method (Goodman, 1964).

Bonett and Price (2012) presented an adjusted Wald interval for matched pairs and 2×2 tables, which introduces a system of weights for those cells involved in the calculation. Alternatively, it is possible to define a confidence interval for the NPS by implementing iterative procedures based on various score tests, such as those based on the original proposal of Wilson (1927), the interactive score method introduced by Tango (1998), which is itself a modification of the test introduced by Agresti and Min (2005) or the May and Johnson (1997) score method. In conclusion, Rocks advises against the use of the Wald and Goodman methods, as they perform poorly. On the contrary, he states that the adjusted Wald method and the iterative score method perform very well, guaranteeing good levels of coverage.

In the paper of Capecchi and Piccolo (2017), the authors search for the distribution of NPS based on a convenient structure of the response patterns. They assume a parametric mixture for the responses and verify the behavior of NPS over the parameter space. From a statistical point of view, they consider NPS index as an estimate of the mean value of a discrete random variable whose probabilities are generated by a distribution expressing the graduated opinions of a sample of respondents on an ordinal scale. In particular, they assume that ordinal responses of the customer judgments/opinions are generated by a CUB (Combination of discrete Uniform and shifted Binomial) model as in Piccolo (2003) and D'Elia and Piccolo (2005). They show that infinitely many CUB models refer to the same NPS and that the uncertainty always present in human decisions as well as the heterogeneity of the respondents may largely affect the NPS value.

Rocks and Capecchi and Piccolo papers represent a significant proposal in which some statistical properties of the NPS index are investigated. This certainly leads to a more accurate description of the index itself but does not overcome all of its criticisms. Our proposal stands alongside that of the cited authors with the aim of investigating, through appropriate statistical procedures, the composition of the index so that companies can implement the appropriate corrective/improvement actions.

3. Critical comparison of similar NPSs

As we have already stated, the same NPSs can represent (very) different situations.

Figure 3 displays all the possible values assumed by the NPS index (from -1 to $+1$) corresponding to all possible numbers of detractors (from no detractors to all respondents being detractors). As we can see from Figure 3, different compositions of the score can give the same result. For example, an NPS of 0.3 can be achieved with detractor percentages from 0% to slightly less than 40% . This raises the question of whether it is reasonable to compare companies with the same NPS while ignoring the percentages of promoters and detractors. More specifically, what conclusions can we draw from the comparison of two (possibly similar) scores for the same company at different points in time, without considering the evolution of these percentages?

This section focuses on comparing two NPSs for the same company at two different points in time, t_1 and t_2 .

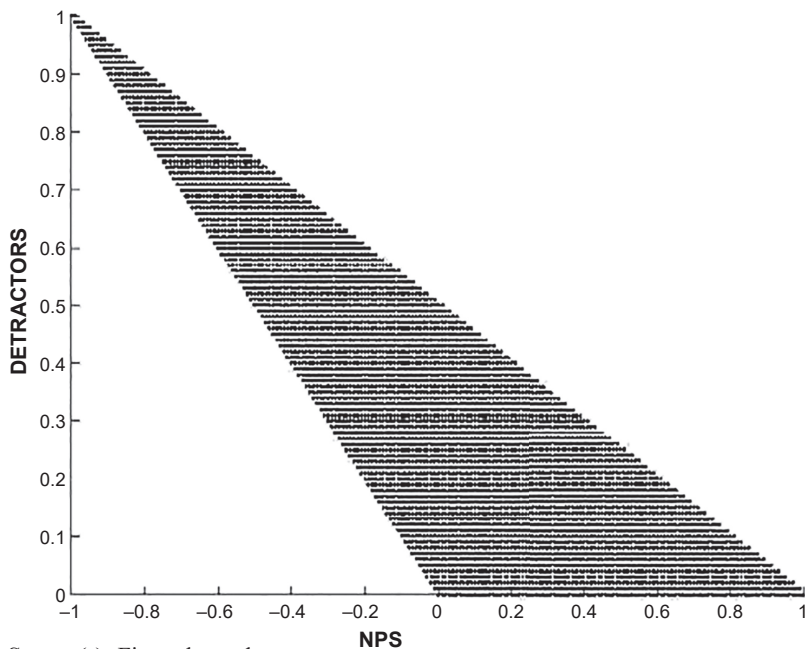
3.1 Composition of the NPS

We consider a company with ratings from 100 customers and their NPS in 2 consecutive years, $Year_1$ and $Year_2$. Consider the situations described in Tables 1 and 2.

Note that the company described in Table 1 has the same NPS in both years in this case:

$$NPS_{Year_1} = NPS_{Year_2} = \frac{50 - 20}{100} = 0.3.$$

The composition of detractors, promoters and passive customers is also the same in both years. Each customer confirms their opinions over time.



Source(s): Figure by authors

Figure 3.
NPS versus all possible
proportions of
detractors

The company described in Table 2 has the same NPS score in both years in this case:

$$NPS_{Year_1} = \frac{72 - 22}{100} = 0.5,$$

$$NPS_{Year_2} = \frac{52 - 2}{100} = 0.5.$$

198

However, the composition is quite different in each year. The customers that are detractors in Year₁ become passive in Year₂, which is good for the company, but the 28% of customers who are promoters in Year₁ shift to passive customers in Year₂, which is not so good for the company. Obviously, the situation in Table 2 is much more realistic than that represented by Table 1.

These two examples highlight the indiscriminate use of the index without evaluating its composition. However, one may reply that the absolute number of promoters and detractors in the two years appears quite different, although it is the same 100 customers. This means that there is some signal that something has changed over time.

In particular, let us consider Tables 3 and 4.

Note that Table 3 has the same marginals as in Table 2 and equal NPS in Year₁ and Year₂ (0.5). However, only about 9% of the customers who are detractors in Year₁ confirm their opinion in Year₂; the remainder are split between passive customers and promoters in Year₂.

Table 1.
Equal NPSs in Year₁
and Year₂, identical
composition of the NPS
index components

Year ₂ Year ₁	Detr	Pass	Prom	Total
Detr	20	0	0	20
Pass	0	30	0	30
Prom	0	0	50	50
Total	20	30	50	100

Source(s): Table by authors

Table 2.
Equal NPSs in Year₁
and Year₂, different
composition of the NPS
index components

Year ₂ Year ₁	Detr	Pass	Prom	Total
Detr	0	22	0	22
Pass	2	4	0	6
Prom	0	20	52	72
Total	2	46	52	100

Source(s): Table by authors

Table 3.
Same marginals as in
Table 2, but with
different compositions
of the NPS index
components

Year ₂ Year ₁	Detr	Pass	Prom	Total
Detr	2	10	10	22
Pass	0	6	0	6
Prom	0	30	42	72
Total	2	46	52	100

Source(s): Table by authors

This is a very good result for the company! Looking only at the NPS value, this result is not detected and, in particular, considering just the NPS values across the years does not highlight the evolution of customers in Tables 2 and 3.

Table 4 presents a situation in which the company has two similar NPSs in the two years (0.52 in Year₁ and 0.5 in Year₂). Note that 100% of the detractors in the first year move and become promoters in the next year. The 31% of promoters in Year₂ change their evaluation in Year₂. Again, these changes in customers' opinions of the company do not emerge from a simple observation of the NPS.

The situations highlighted in Tables 3 and 4 are clearly borderline case studies. In reality, it will be quite difficult to find a detractor of a company that becomes a promoter from one year to the next. The objective of these considerations is a mathematical study of the NPS index, and these situations illustrate the limitations of the indicator itself.

In this subsection, we have highlighted the different compositions of detractors, passive customers and promoters that can produce similar NPSs from a descriptive point of view. In the next section, we consider the situation from an inferential perspective.

4. Methodology

A statistical validation of equal NPSs at two different points in time can be achieved by looking at the marginal data in the tables presented in the previous subsection. Equal NPSs with a "similar" composition of components should have a two-way contingency table that satisfies the *marginal homogeneity* hypothesis.

4.1 Marginal homogeneity

Let $(NPS_{Year_1}, NPS_{Year_2})$ denote the two responses of a randomly selected matched set. With three response categories, a contingency table with 3×3 cells summarizes the possible outcomes.

Let $\mathbf{j} = (j_1, j_2)$ denote the cell containing $NPS_{Year_t} = j_t, t = 1, 2$. Let $\pi_{\mathbf{j}} = P(NPS_{Year_t} = j_t, t = 1, 2)$ be the joint distribution of $(NPS_{Year_1}, NPS_{Year_2})$. Then,

$$P(NPS_{Year_t} = j) = \pi_{+j},$$

where the subscript j is in position t and the subscript $+$ denotes the sum over that index.

Note that $\{P(NPS_{Year_t} = j), j = 1, 2, 3\}$ is the marginal distribution for NPS_{Year_t} [3]. This two-way table satisfies marginal homogeneity if

$$P(NPS_{Year_1} = j) = P(NPS_{Year_2} = j), \text{ for } j = 1, 2, 3.$$

Tests of marginal homogeneity have been studied for binary contingency tables and extended to larger tables (Agresti, 2013, Ch. 11). Such tests can differentiate between nominal and ordinal variables.

Year ₂ Year ₁	Detr	Pass	Prom	Total
Detr	0	0	20	20
Pass	6	2	0	8
Prom	14	8	50	72
Total	20	10	70	100

Source(s): Table by authors

Table 4.
Similar NPSs in Year₁
and Year₂, but different
compositions of the
NPS index components

In our case of ordinal variables, we compare the marginals using a cumulative marginal logit model:

$$\text{logit}[P(NPS_{Year_t} \leq j | x_t)] = \alpha_j + \beta \cdot x_t \text{ for } t = 1, 2, j = 1, 2, \quad (1)$$

where $x_1 = 0, x_2 = 1$ and $\text{logit}[P(NPS_{Year_t} \leq j | x_t)]$ for $t = 1, 2$ and $j = 1, 2, 3$ denotes the so-called *cumulative logit*:

$$\text{logit}[P(NPS_{Year_t} \leq j | x_t)] = \ln \frac{P(NPS_{Year_t} \leq j | x_t)}{1 - P(NPS_{Year_t} \leq j | x_t)}.$$

Each cumulative logit uses all three response categories. Note that this model simultaneously uses two cumulative logits for $NPS_{Year_t}, t = 1, 2$. Following Eq. (1), each cumulative logit has its own intercept α_j . The α_j are increasing in j , because $P(NPS_{Year_t} \leq j)$ increases in j and the logit is an increasing function of $P(NPS_{Year_t} \leq j)$.

Usually, the α_j intercepts are not of interest except for computing response probabilities. The parameter estimates yield estimated logits and hence estimates of $P(NPS_{Year_t} \leq j | x_t)$ or $P(NPS_{Year_t} > j | x_t)$. It is worthwhile to note that this model gives stochastically ordered marginal distributions, with $\beta > 0$ indicating that NPS_{Year_1} tends to be higher than NPS_{Year_2} . Marginal homogeneity corresponds to $\beta = 0$. The further role of the β parameter will be highlighted in the next subsection.

Maximum likelihood (ML) fitting of this model is not straightforward (model fitting treats $(NPS_{Year_1}, NPS_{Year_2})$ as dependent, Agresti 2013, Ch. 12), but can it be done using the R statistical software (R Core Team, 2019) through the specialized `mph.fit` function developed by Joseph Lang at the University of Iowa, which is contained in the `hmmm` package (Colombi et al., 2014). The ML marginal fitting method makes no assumptions about the model that describes the joint distribution of π_j . Thus, when the model holds, the ML estimate of parameters is consistent regardless of the dependence structure for that distribution.

The marginal homogeneity model (H_0 : *marginal homogeneity*, $\beta = 0$; H_1 : $\bar{H}_0, \beta \neq 0$) is validated through the likelihood ratio test G^2 , which compares the model under investigation (marginal homogeneity) with the saturated (unconstrained) one. Under the null hypothesis, the test statistic G^2 follows the χ^2 distribution with degrees of freedom, df , equal to the difference between the free parameters in the two models (the saturated model and the tested model). We reject the hypothesis that the selected model provides a good representation of the dataset when the *p-value* is less than some critical value (usually 0.05).

4.2 Cumulative odds ratio

The cumulative marginal logit model assumes a proportional odds structure, which means that it has the same effect β for each logit; indeed, this model satisfies Eq. (2):

$$\text{logit}[P(NPS_{Year_2} \leq j)] - \text{logit}[P(NPS_{Year_1} \leq j)] = \beta. \quad (2)$$

Therefore, the same proportionality constant applies to each logit. Furthermore,

$$\begin{aligned} \ln \frac{P(NPS_{Year_2} \leq j)}{P(NPS_{Year_2} > j)} - \ln \frac{P(NPS_{Year_1} \leq j)}{P(NPS_{Year_1} > j)} &= \beta, \\ \ln \left[\frac{P(NPS_{Year_2} \leq j)}{P(NPS_{Year_2} > j)} \cdot \frac{P(NPS_{Year_1} > j)}{P(NPS_{Year_1} \leq j)} \right] &= \beta, \end{aligned}$$

$$\frac{P(NPS_{Year_2} \leq j)}{P(NPS_{Year_2} > j)} \cdot \frac{P(NPS_{Year_1} > j)}{P(NPS_{Year_1} \leq j)} = \exp(\beta), \tag{3}$$

$$\frac{P(NPS_{Year_2} \leq j)}{P(NPS_{Year_2} > j)} = \exp(\beta) \cdot \frac{P(NPS_{Year_1} \leq j)}{P(NPS_{Year_1} > j)}, \text{ for } j = 1, 2. \tag{4}$$

Note that, in the above formulas, we have omitted the references to $x_t, t = 1, 2$, to simplify the notation. Indeed, from Eq. (4), the odds of the outcome $NPS_{Year_2} \leq j$ is $\exp(\beta)$ times the odds of $NPS_{Year_1} \leq j$ for $j = 1, 2$. This is why the cumulative marginal logit model is often called the “proportional odds model” (McCullagh, 1980). Note that an odds ratio of cumulative probabilities, as given by $\exp(\beta)$ in Eq. (3), is called a *cumulative odds ratio*.

We have already stated that, in the cumulative marginal logit model, the marginal homogeneity corresponds to $\beta = 0$. This implies that:

$$\frac{P(NPS_{Year_2} \leq j)}{P(NPS_{Year_2} > j)} = \frac{P(NPS_{Year_1} \leq j)}{P(NPS_{Year_1} > j)} \text{ for } j = 1, 2,$$

which means that the cumulative odds ratio $\exp(\beta)$ is equal to 1. In cases where the hypothesis of marginal homogeneity is rejected, the cumulative odds ratio becomes an interesting tool for measuring the proportionality between the odds.

5. Results

Applying the marginal homogeneity model to the tables presented in Subsection 3.1, we obtain the results in Table 5. Obviously, Table 1 represents the marginal homogeneity situation.

Examining the marginals of the three tables considered in Table 5, the decisions according to the marginal homogeneity tests (Reject H_0 , Reject H_0 and Do not reject H_0 , respectively) are quite obvious for all three cases. Note that all three tables have broadly similar NPSs over time. Performing this kind of statistical test brings out details on the composition of the index that are hidden when looking at only a single number. Furthermore, it is worthwhile considering the situation described in Table 6.

	G^2	p - value	Decision
Table 2	47.2288	0.0000	Reject H_0
Table 3	55.5600	0.0000	Reject H_0
Table 4	0.3133	0.8550	Do not reject H_0

Source(s): Table by authors

Table 5.
Marginal homogeneity model: H_0 : *marginal homogeneity*, H_1 : \bar{H}_0

Year ₂ Year ₁	Detr	Pass	Prom	Total
Detr	18	3	1	22
Pass	0	6	0	6
Prom	2	1	69	72
Total	20	10	70	100

Source(s): Table by authors

Table 6.
Equal NPSs in Year₁ and Year₂, as in Tables 2 and 3, and similar compositions of the NPS index components

In this case, $NPS_{Year_1} = NPS_{Year_2} = 0.5$ once again. The two indices have apparently similar compositions over time. In fact, 4 of the 22 detractors and 3 of the 72 promoters in *Year*₁ change their opinions. The marginal homogeneity model applied to this table gives the following results: $G^2 = 5.5460$ with $p\text{-value} = 0.0625$. Thus, with the usual benchmark level of significance, we will not reject the marginal homogeneity hypothesis, but this does not happen with higher levels of significance (i.e. 10%). This situation highlights that even slight changes in opinion of the detractors/promoters give statistically significant consequences.

As we already mentioned, in cases where the hypothesis of the marginal homogeneity is rejected, the cumulative odds ratio becomes an interesting tool for measuring the proportionality between the odds. Table 7 reports the estimated cumulative odds ratio of the tables for which the hypothesis of marginal homogeneity was rejected.

The interpretation of the estimated cumulative odds ratio comparing marginals is $\exp(\hat{\beta})$ as highlighted in Eqs. (3) and (4). This means that in Table 2, the estimated odds of the response “detractor” in *Year*₂ for a randomly selected subject are $e^{0.0975} = 1.1$ times the estimated odds of the response “detractor” in *Year*₁ for another randomly selected subject. Additionally, the estimated odds of the response “detractor” or “passive” in *Year*₂ for a randomly selected subject are 1.1 times the estimated odds of the response “detractor” or “passive” in *Year*₁ for another randomly selected subject. Considering Table 3, the estimated odds of the response “detractor” in *Year*₂ for a randomly selected subject are $e^{0.6437} = 1.9$ times the estimated odds of the response “detractor” in *Year*₁ for another randomly selected subject. The estimated odds of the response “detractor” or “passive” in *Year*₂ for a randomly selected subject are 1.9 times the estimated odds of the response “detractor” or “passive” in *Year*₁ for another randomly selected subject. At this point, it is worth comparing the situations represented in Tables 2 and 3. They present the same values of the NPS index in the two years being considered. Analysis of this single number could suggest similar situations. We have already highlighted how the composition of the components of the index differs in the two situations. In particular, in Table 2, the second year shows an improvement in “detractors” and a worsening in “promoters.” In Table 3, however, the situation improves considerably from one year to the next. This diversity between the two tables emerges with the marginal homogeneity test, which rejects the hypothesis of homogeneity. Now, the fact that the two tables represent different cases begins to be evident. This evidence becomes even stronger with the use of the cumulative odds ratio. In Table 2, the possibility of having been in the same condition (detractor or passive) from one year to the other varies, but it is much less than in Table 3, where, instead, it almost doubles.

Other situations are worth investigating as well. For example, consider Table 4. As already mentioned, the marginal homogeneity test indicates that Table 4 presents the homogeneous marginals as is evident to the naked eye. In Table 4, therefore, one would expect an estimate of the cumulative odds ratio equal to 1. In this case, instead, it is equal to $e^{1.6797} = 5.4$, a value very far from 1! This apparently unexpected result is actually justified by the particular situation represented by Table 4, where there are so-called “compensations” in the marginal distributions. Therefore, an investigation of the homogeneity of the marginal distributions would not, in this case, have been sufficient to highlight the different compositions of the index in the two periods considered.

		$\hat{\beta}$	$\exp(\hat{\beta})$
Table 7. Estimated cumulative odds ratio, $\exp(\hat{\beta})$	Table 2	0.0975	1.1
	Table 3	0.6437	1.9
Source(s): Table by authors			

6. Discussion, implications and further research

6.1 Discussion

Many methods in the literature, as cited in the Introduction section, that measure customer satisfaction use statistical techniques to obtain results on which to base business management strategies. For example, Structural Equation Modeling (SEM) is usually the technique for finding the customer satisfaction level and validating the causal relationship between customer satisfaction and its antecedents and consequences. This technique is, therefore, used to validate different types of customer satisfaction indices. The objective of SERVQUAL methodology is usually to develop the best instrument for measuring customer satisfaction and SEM; Factor Analysis or Multiple Regression analysis are usually used for choosing and validating the best service quality constructs among the proposed ones. Furthermore, the MUSA method follows the principles of ordinal regression analysis under constraints.

It should be noted that the literature that has dealt with NPS has mainly focused on highlighting the weaknesses of the indicator. Solutions are suggested to overcome these weaknesses but often no mathematical-statistical models are implemented to verify the validity of the proposed solutions, e.g. changing the scale. Other works provide indications as to how management should behave, e.g. additional surveys (see [Subsection 1.1](#)). An innovative methodological proposal is that of [Rocks \(2016\)](#) who, by defining the probabilistic context of the index, determines the construction of confidence intervals around the index value estimate (see [Section 2](#)). In addition, [Capecchi and Piccolo \(2017\)](#) search for the distribution of NPS based on a convenient structure of the response patterns. Furthermore, fuzzy set Qualitative Comparative Analysis (fsQCA) has been used to analyze the relationship between customer satisfaction and loyalty measured by the NPS and dependent variables as gastronomy, cleanliness and room comfort and satisfaction expressed by clients in the area of reception in the hotel industry by [Baquero \(2022\)](#).

Our proposal stands as a bridge between the pure management approach and the application of statistical models. The intent of our proposal is to offer a statistical tool known in the literature and easy to use and read in order to facilitate the company management in the correct reading and interpretation of the NPS. We are inspired by Deming's TQM philosophy. His PDCA cycle in our proposal can be interpreted as follows (see [Johnson, 2002](#); [Taufik, 2020](#)):

- (1) **Plan**: plan the change. Plan consists of setting goals and strategies to achieve specific results.
- (2) **Do**: test the change.
- (3) **Check**: analyze the results and identify learnings.
- (4) **Act**: take action based on what you learned in the check step.

Our proposal is summarized in [Figure 4](#).

[Figure 4](#) can be interpreted as follows.

Plan: the company starts computing NPS_{t_1} and sets goals and strategy to achieve in the reference period of time.

Do: the company computes NPS_{t_2} and compares to NPS_{t_1} . If $NPS_{t_1} \neq NPS_{t_2}$ enter in the **check** phase and evaluate the future actions. **Act**: if needed to achieve the business growth goals. If $NPS_{t_1} = NPS_{t_2}$, the same NPS value can actually represent very different situations and then enter in the **check** phase, first, performing a marginal homogeneity test of H_0 : *marginal homogeneity* vs $H_1 : \bar{H}_0$ and calculate the estimated cumulative odds ratio.

- (1) If H_0 is not rejected and the **estimated cumulative odds ratio is equal to 1**, then we can consider the margins to be homogeneous, i.e. [Table 1](#). In this case,

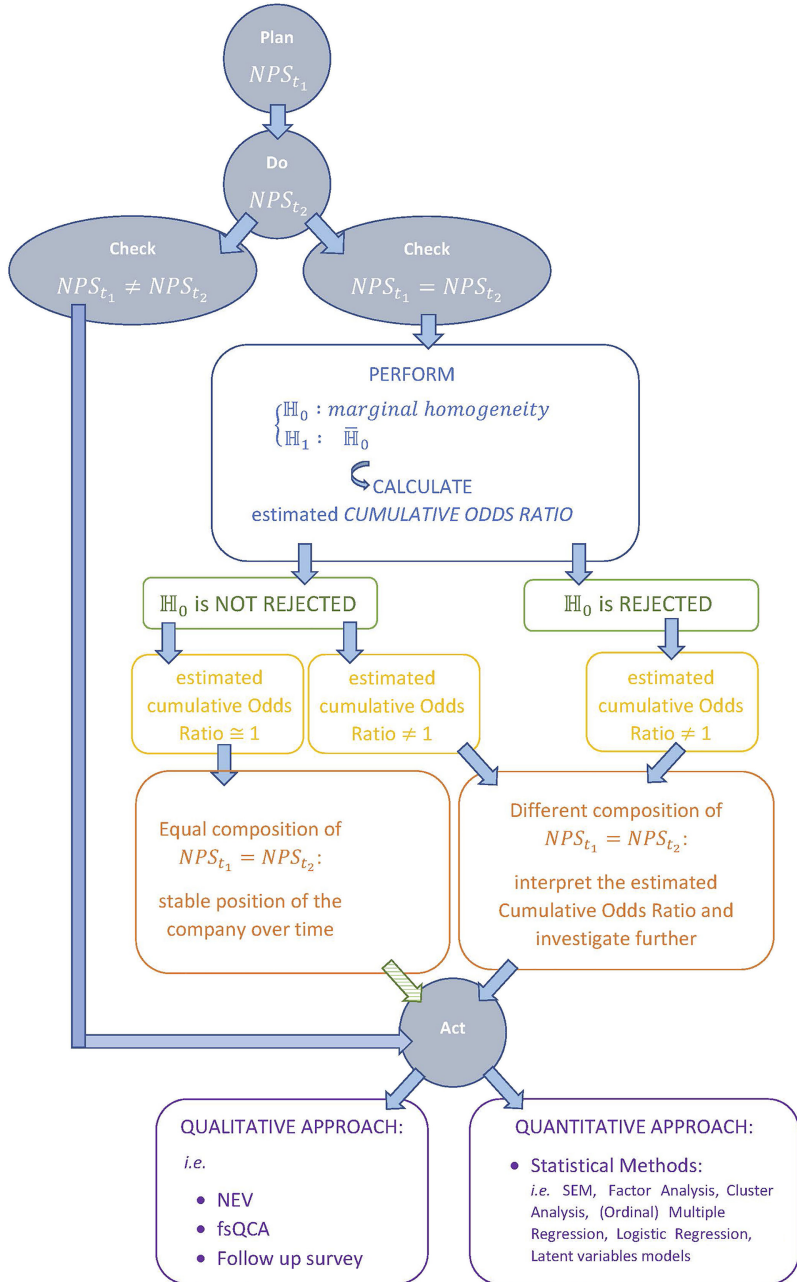


Figure 4.
Decision algorithm

Source(s): Figure by authors

$NPS_{t_1} = NPS_{t_2}$ indicates an equal composition between the two indices. This scenario describes the situation in which the company has maintained a stable position over time with regard to the “loyalty” of its clients. There has been neither deterioration nor improvement.

Act: in this case, the company, having assessed the degree of dynamism of the market in which it operates, may decide to improve its market position carrying out ad hoc surveys among its customers to find out which aspects to improve or maintain its established position in the market.

- (2) If H_0 is not rejected and the **estimated cumulative odds ratio is far from 1**, then we can consider the marginals to be homogeneous because of compensation. In this case, $NPS_{t_1} = NPS_{t_2}$ does not indicate an equal composition between the two indices, i.e. [Table 4](#). To check how the situation evolved between NPS_{t_1} and NPS_{t_2} , consider the estimated cumulative odds ratios and judge how the compositions have changed in the considered situations. This represents the most ambiguous case. In fact, the first information given by the statistical survey would lead to conclusions that are the opposite of those when the survey is complete. This is the case that best highlights the criticality of the NPS index. Therefore, it is necessary to have further statistical instruments to confirm (or not) the information apparently provided by the index itself.

Act: in this case, the company has to investigate further, by choosing whether to investigate according to a qualitative or quantitative approach, taking advantage of the different methodologies existing in the literature.

- (3) If H_0 is rejected and the **estimated cumulative odds ratio is far from 1**, then we can consider the margins not to be homogeneous. In this case, $NPS_{t_1} = NPS_{t_2}$ but the composition of the two indices differs, i.e. [Tables 2 and 3](#). To check how the situation has evolved between NPS_{t_1} and NPS_{t_2} , consider the estimated cumulative odds ratios and judge how the compositions have changed in the considered situations. This scenario represents the most extreme theoretical situation in which the company must understand how its position has changed, for better or worse, in order to implement any corrective actions.

Act: in this case, the company that wants to improve its position in the market has to investigate further. In particular, it should carry out ad hoc surveys among its customers in order to understand the reasons why customers responded favorably/unfavorably. In addition to the single question used for the construction of the NPS, other questions could be added that aim to clarify the reasons why the customer gave a certain grade/score ([Rajasekaran and Dinesh, 2018](#)). In this sense, one could also proceed with the Net Emotional Value (NEV), i.e. try to analyze the customer's experience through the study of his or her emotions, thus creating a greater connection with the company itself ([Achmad et al., 2020](#)). Basically, companies that find themselves in this position are necessarily faced with an obvious situation of dissatisfaction on the part of their customers. A valid solution is to choose whether to investigate this loss of consensus on the part of their customers according to a qualitative or quantitative approach, taking advantage of the different methodologies existing in the literature.

The aim of this work has been to draw attention to the indiscriminate use of the NPS index. In particular, we highlighted how the same NPS value can actually represent very different situations. We have proposed a statistical validation of the use of this index by suggesting

structures that are already known in the literature and that can easily support the analysis of the NPS index.

6.2 Theoretical and practical implications

Following the research line traced by the study, various theoretical and practical implications can be derived.

Referring to theory, first this study contributes to the current literature by adopting a statistical approach to determine whether or not identical values of the index can be considered similar based on statistical assumptions, adding novel knowledge in an under-researched topic in the NPS literature.

Furthermore, it has already been pointed out that the management of a company often makes decisions based on subjective threshold values of the NPS index. Our proposal would make it possible to statistically validate the choice of these threshold values in a more objective manner. Our algorithm also allows for temporal comparisons of the index and can thus support PDCA actions to be carried out over time. By implications, the theory allows individuals and organizations to plan and continually improve themselves, their relationships, processes, products and services.

Another insight for scholars of NPS users is that a successful and vigorous implementation of our algorithm improves positively the conscious knowledge of the proper customer loyalty. In particular, note that the technical implementation of our algorithm is feasible with any basic statistical software, e.g. the free R software.

Furthermore, we think our proposal can help a company improve its quality management. We are certainly not able to provide data on what causes dissatisfaction. We can, however, indicate that its NPS index has changed composition from one time instant to the next and, for example, point out to the company that some of its promoters have become passive. Many business firms are channeling more efforts to retain existing customers rather than acquiring new customers since the cost of acquiring new customers is greater than retaining the existing ones. This information will enable the company to activate all the procedures, which it is able to manage, in the Act phase of the PDCA cycle to achieve the goal of improving its next NPS.

In confirmation of what has already been presented, it is worth noting that large companies (e.g. HP and Sky) have already implemented this good practice of combining the “single question” survey of the NPS with a questionnaire investigating the reasons supporting the summary judgment made by the NPS itself. In our opinion, these companies have already incorporated, according to Deming’s philosophy, the need to capture customer satisfaction/dissatisfaction reasons.

6.3 Limitations and further research

There has been a great deal of debate in the literature on the erroneous and illusory use of the NPS: there is no scientific confirmation of the link between the value of the index and growth in customer loyalty. Some scholars believe that ignoring the large proportion of neutrals is a big mistake. Being passive does not necessarily mean having a neutral stance; in fact, they may be more likely to assimilate with detractors in terms of searching for a better buying experience. There is also no evidence that the value of the NPS is a good predictor of future sales growth. Finally, the NPS is not even reliable in measuring the growth/decline of a company over time (Mecredy *et al.*, 2018; Fisher and Kordupleski, 2019).

This work does not pretend to be exhaustive of all the criticisms that have emerged regarding the NPS. Instead, we have tried to highlight the usefulness of NPS users possessing the basic statistical knowledge that is necessary to be able to use tools that make the index

itself more effective, according to our proposal. Furthermore, it is noted that the index becomes much more reliable when a long historical series of data is available, allowing a longitudinal reading of the company's performance over time.

Our proposal also has limitations, of course. The statistical methodology we have proposed is in fact applicable to those indicators that, like the NPS, can be measured at two different time instants in a contingency table. However, indicators that are direct competitors of the NPS (NCSI, ACSI and EPSI) could be studied in this way if – for example, one could detect the responses of customers with respect to the individual levels of the items generating the indicators themselves.

The authors identified these further strands of research as elements to be worked on:

- (1) Revision of the scoring scale: testing the appropriateness of moving from an 11-point scale to a smaller number of attributable scores. The proposal is to switch to scales typical of psycho-sociological disciplines, in line with what has already been outlined in the literature, see, among others, [Schneider et al. \(2008\)](#),
- (2) Revising the division of the 11-scale into the 3 classifications. The proposal is to seek a better distribution of the scores attributed to the three classes (promoters, passives and detractors) leading to a greater correspondence between score and classification of the client providing the answer as suggested by i.e. [Kristensen and Eskildsen \(2014\)](#),
- (3) Transformation of the index through the introduction of an appropriate system of weights that highlights the contribution of the individual components that cannot be considered equivalent for the purposes of index composition. In a first analysis, the weights could reflect state transition probabilities;
- (4) Propose the methodology in finer tables, with the index responses directly in the 0–10 scale. In this way, the method can also be extended to validations of other indices by constructing contingency tables corresponding to the levels of the indices. This could potentially allow the sensitivity of the indices to shifts in respondents' choices to be assessed.

Our paper confirms that although a single number summarizes and communicates a complex situation very quickly, especially with audiences that are not in a position to engage in a very technical discussion, it is ambiguous and unreliable if not accompanied by other statistical tools.

Notes

1. <https://measuringu.com/nps-ux/>
2. <https://www.researchworld.com/stop-using-nps-net-promoter-score-but-please-save-the-question/>
3. Note that $P(NPS_{Year_t} = 1) = P(NPS_{Year_t} = 'Detr')$; $P(NPS_{Year_t} = 2) = P(NPS_{Year_t} = 'Pass')$; $P(NPS_{Year_t} = 3) = P(NPS_{Year_t} = 'Prom')$ for $t = 1, 2$.

References

- Achmad, S.A., Anggina, P. and Rudito, P. (2020), "Strategic planning customer experience using predictive analysis Indihome PT Telkom", *IPTEK Journal of Proceedings Series*, Vol. 1, pp. 457-468.
- Agresti, A. (2013), *Categorical Data Analysis*, 3rd ed., Wiley, Hoboken, New Jersey.
- Agresti, A. and Coull, B.A. (1998), "Approximate is better than "exact" for interval estimation of binomial proportions", *The American Statistician*, Vol. 52, pp. 119-126.

- Agresti, A. and Min, Y. (2005), "Simple improved confidence intervals for comparing matched proportions", *Statistics in Medicine*, Vol. 24, pp. 729-740.
- Arora, P. and Narula, S. (2018), "Linkages between service quality, customer satisfaction and customer loyalty: a literature review", *IUP Journal of Marketing Management*, Vol. 17 No. 4, p. 30.
- Baehre, S., O'Dwyer, M., O'Malley, L. and Lee, N. (2022), "The use of net promoter score (NPS) to predict sales growth: insights from an empirical investigation", *Journal of the Academy of Marketing Science*, Vol. 50, pp. 67-84.
- Baquero, A. (2022), "Net promoter score (NPS) and customer satisfaction: relationship and efficient management", *Sustainability*, Vol. 14 No. 4, p. 2011.
- Bonett, D.G. and Price, R.M. (2012), "Adjusted Wald confidence interval for a difference of binomial proportions based on paired data", *Journal of Educational and Behavioral Statistics*, Vol. 37, pp. 479-488.
- Capecchi, S. and Piccolo, D. (2017), "The distribution of Net Promoter Score in socio-economic surveys", in *Statistics and Data Science, New Challenges, New Generations*, pp. 247-252.
- Colombi, R., Giordano, S. and Cazzaro, M. (2014), "Hmmm: an R package for hierarchical multinomial marginal models", *Journal of Statistical Software*, Vol. 59 No. 11, pp. 1-25.
- D'Elia, A. and Piccolo, D. (2005), "A mixture model for preferences data analysis", *Computational Statistics and Data Analysis*, Vol. 49 No. 3, pp. 917-934.
- de Laplace, P.S. (1812), *Théorie Analytique des Probabilités*, Courcier, Paris, France.
- De Luca, A. (2006), "A logit model with a variable response and predictors on an ordinal scale to measure customer satisfaction", *Quality and Reliability Engineering International*, Vol. 22, pp. 591-602.
- Deming, W.E. (1986), *Out of the Crisis* Cambridge, Massachusetts Institute of Technology, Massachusetts.
- East, R., Romaniuk, J. and Lomax, W. (2011), "The NPS and the ACSI: a critique and an alternative metric", *International Journal of Market Research*, Vol. 53, pp. 2-16.
- Eskildsen, J. and Kristensen, K. (2011), "The accuracy of the net promoter score under different distributional assumptions", *IEEE*, pp. 964-969.
- Fisher, N.I. and Kordupleski, R.E. (2019), "Good and bad market research: a critical review of net promoter score", *Applied Stochastic Models in Business and Industry*, Vol. 35, pp. 138-151.
- Gold, R.Z. (1963), "Text auxiliary to χ^2 tests in a Markov chain", *Annals of Mathematical Statistics*, Vol. 34, pp. 56-74.
- Goodman, L.A. (1964), "Simultaneous confidence intervals for contrasts among multinomial populations", *Annals of Mathematical Statistics*, Vol. 35, pp. 716-725.
- Goodman, L.A. (1965), "On simultaneous confidence intervals for multinomial proportions", *Technometrics*, Vol. 7, pp. 274-254.
- Johnson, C.N. (2002), "The benefits of PDCA", *Quality Progress*, Vol. 35 No. 5, p. 120.
- Kristensen, K. and Eskildsen, J. (2014), "Is the NPS a trustworthy performance measure?", *The TQM Journal*, Vol. 2, pp. 202-214.
- May, W.L. and Johnson, W.D. (1997), "Confidence intervals for differences in correlated binary proportions", *Statistics in Medicine*, Vol. 16, pp. 2127-2136.
- McCullagh, P. (1980), "Regression models for ordinal data", *Journal of the Royal Statistical Society B*, Vol. 42, pp. 109-142.
- McCreedy, P., Wright, M.J. and Feetham, P. (2018), "Are promoters valuable customers? An application of the net promoter scale to predict future customer spend", *Australasian Marketing Journal*, Vol. 26, pp. 3-9.
- Ngo, V.M. (2015), "Measuring customer satisfaction: a literature review", *Proceedings of the 7th International Scientific Conference Finance and Performance of Firms in Science*, Vol. 7, pp. 1637-1654, Education and Practice.

-
- Piccolo, D. (2003), "On the moments of a mixture of uniform and shifted binomial random variables", *Quaderni di Statistica*, Vol. 5 No. 1, pp. 85-104.
- Pingitore, G., Morgan, N.A., Rego, L.L., Gigliotti, A. and Meyers, J. (2007), "The single- question trap", *Marketing Research*, Vol. 19, pp. 9-13.
- R Core Team (2019), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. available at: <https://www.Rproject.org>
- Rajasekaran, N. and Dinesh, N. (2018), "How net promoter score relates to organizational growth", *International Journal of Creative Research Thoughts*, Vol. 6 No. 2, pp. 972-981.
- Reichheld, F.F. (2003), "The one number you need to grow", *Harvard Business Review*, Vol. 12, pp. 46-54.
- Reichheld, F.F. and Markey, R. (2011), *The Ultimate Question 2.0: How Net Promoter Companies Thrive in a Customer-Driven World*, Harvard Business Press, Boston, Massachusetts.
- Rocks, B. (2016), "Interval estimation for the "net promoter score", *The American Statistician*, Vol. 70, pp. 365-372.
- Schneider, D., Berent, M., Thomas, R. and Krosnick, J. (2008), "Measuring customer satisfaction and loyalty: improving the 'Net-Promoter' score", *Poster presented at the Annual Meeting of the American Association for Public Opinion Research*, New Orleans, Louisiana.
- Sharp, B. (2006), "Net promoter score fails the test", *Marketing Research*, Vol. 20, pp. 28-30.
- Tango, T. (1998), "Equivalence test and confidence interval for the difference in proportions for the paired-sample design", *Statistics in Medicine*, Vol. 17, pp. 891-908.
- Taufik, D. (2020), "PDCA cycle method implementation in industries: a systematic", *IJIEM (Indonesian Journal of Industrial Engineering and Management)*, Vol. 1, pp. 157-166.
- Wilson, E.B. (1927), "Probable inference, the law of succession, and statistical inference", *Journal of the American Statistical Association*, Vol. 22, pp. 209-212.
- Worlu, R.E., Adeniji, A.A., Atolagbe, T.M. and Salau, O.P. (2019), "Total quality management (Tqm) as a tool for sustainable customer loyalty in a competitive environment: a critical review", *Academy of Strategic Management Journal*, Vol. 18 No. 3, pp. 1-6.
- Zanella, A. (1998), "A statistical model for the analysis of customer satisfaction: some theoretical and simulation results", *Total Quality Management*, Vol. 9 No. 7, pp. 599-609.

Corresponding author

Manuela Cazzaro can be contacted at: manuela.cazzaro@unimib.it

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com